SUPPLEMENTARY MATERIALS AND FIGURES

"Resection of individually identified high-rate high-frequency oscillations region is associated with favorable outcome in neocortical epilepsy"

Cho JR, Hong SB, Koo DL, Joo EY, Seo DW, Hong SC, Jiruska P

Human labeling and gold standard dataset

Evaluation, optimization and validation of automatic HFO detector require gold standard events reliably identified by human reader. To evaluate the performance of automated HFO detection method, we compared the detection results with a human reviewer experienced in HFO (Cho JR). One-minute of interictal data from two channels (one from SOZ channels and another from non-SOZ channels) per each patient were extracted. The orders of extracted data were randomly shuffled and patient and channel information were concealed. These data, along with band-passed signals (one for R range and another for FR range), were simultaneously shown in the custom-built graphic user interface in MATLAB to facilitate visual marking. We adopted an approach recommended by the Montreal group (for guidelines how to visually identify HFO, see Zijlmans *et al.*, 2012): a ripple was marked if an event is clearly distinguishable from background in the R range filtered signal but not in FR filtered signal, and a fast ripple was marked if an event is evident only in the FR range filtered signal. Only events with at least three consecutive oscillations were considered. Signal epochs (duration 500 ms) containing labeled events were extracted and stored for further analysis. These events were later decomposed with Morlet wavelet (Supplementary Fig. 2 and 3). The wavelet decomposition map helped the reviewer to select the final set of visually validated HFO events: valid events have to show spectrally, temporally isolated "blob" in either R or FR range the decomposition map. False events were carefully rejected. In total, 514 R events and 367 FR events were selected as gold standard events.

Performance evaluation of the first stage detector

Performance of the first stage of the detector was optimized for amplitude threshold values ranging from 0.5 to 5.5 SD. Optimal amplitude threshold value (3 SD) was determined by analysis of receiver-operating characteristics (ROC) curves (Supplementary Fig. 4). Two measures were used to quantify the performance of the detector. To determine the sensitivity we used accuracy measure. Accuracy was defined as the percentage of true positive detections compared to a ground truth data set. Positive predictive value (PPV) was defined as the number of true positive detections divided by the sum of the number of true positive and false positive detections. Performance of the first stage of the optimized detector (amplitude threshold of 3 SD) was characterized by the accuracy 97.28% for R and 93.46% for FR. PPV was 56.75% for R and 32.23% for FR (Supplementary table 1). These results suggest that the first-stage detector achieves high sensitivity but suffers from large number of false positive detections, since we deliberately used low amplitude thresholds.

Identification of common false positive detections

Consistent with previous reports (Crepon *et al.*, 2008; Benar *et al.*, 2010; Blanco *et al.*, 2010), the first-stage automated detector generated high number of false-positive detections. Comparison with gold standard datasets allowed us to identify three of the most common types of false positive detections:

1) ***Type 1 -*** high-frequency transients (Supplementary Fig. 5) due to high-pass filtering of

sharp components of interictal epileptiform discharges, i.e. discharges with no visible superimposed fast oscillations in the unfiltered signal and/or with absence of isolated fast activity ("blobs") in time-frequency decomposition plane.

2) ***Type 2 -*** harmonics of low-frequency, non-sinusoidal signals (Supplementary Fig. 6)
3) ***Type 3 -*** transient events with amplitudes larger than the global background (10 minute) but not significantly different from the local (250 ms preceding and following detections) activity (Supplementary Fig. 7).

These types of detected events have morphological and spectral features that can be distinguished from pathologic HFO. Therefore, three post-processing steps (see *Methods*) were implemented to reject these types of false detections. Overall, the power band ratio was effective in eliminating type 1 and type 2 while local background comparison was useful to reject type 3 false detections.

Optimization and evaluation of the post-processing stages

Implementation of the second-stage post-processing slightly decreased the accuracy (90.46% for R and 84.20% for FR) and substantially improved PPV (90.29% for R and 81.32% for FR). Although second-stage post-processing steps incorrectly rejected some true HFO events, its effect is minimal and the overall performance was greatly improved due to noticeable enhancement in specificity for both HFO features. This effect is illustrated in the Supplementary Figure 4. Overall, the performance of our automated detection method was comparable to that of manual visual marking.

Quantitative evaluation of detected HFO events

For detected HFO events, we quantified three features: 1) peak frequency, 2) peak amplitude, and 3) duration. To calculate peak frequency, power spectral density (PSD) was computed on R- or FR band-pass filtered signal using a multi-taper method. Frequency value at maximum power was declared as peak frequency. Peak amplitude is found from the maximum value of signal envelope (instantaneous amplitude) of detected events. Duration was automatically calculated as the time in ms (converted from the number of samples) of detected segments.

The mean peak amplitude of R was 48.5 ± 51.9 μV (median 36.18 μV), the mean duration 51.8 ± 27.9 ms (median 43.5 ms), and the peak frequency 90.1 ± 19.4 Hz (median 85.9 Hz). Almost 85% (48,120/56,844) of R had peak frequency less than 100 Hz. The mean peak amplitude of FR was 26.6 ± 35.9 μV (median 12. 38 μV), the mean duration 19.5 ± 9.7 ms (median 16.5 ms), and the peak frequency 255.8 ± 56.3 Hz (median 234.4 Hz). Histograms are illustrated in Supplementary Figure 8.

References

Benar CG, Chauviere L, Bartolomei F, Wendling F. Pitfalls of high-pass filtering for detecting epileptic oscillations: a technical note on "false" ripples. Clin Neurophysiol 2010; 121: 301-310.

Blanco JA, Stead M, Krieger A, Viventi J, Marsh WR, Lee KH, Worrell GA, Litt B. Unsupervised classification of high-frequency oscillations in human neocortical epilepsy and control patients. J Neurophysiol 2010; 104: 2900-2912.

Crepon B, Navarro V, Hasboun D, Clemenceau S, Martinerie J, Baulac M, Adam C, Le Van Quyen M. Mapping interictal oscillations greater than 200 Hz with intracranial macroelectrodes in human epilepsy. Brain 2010: 133: 33-45

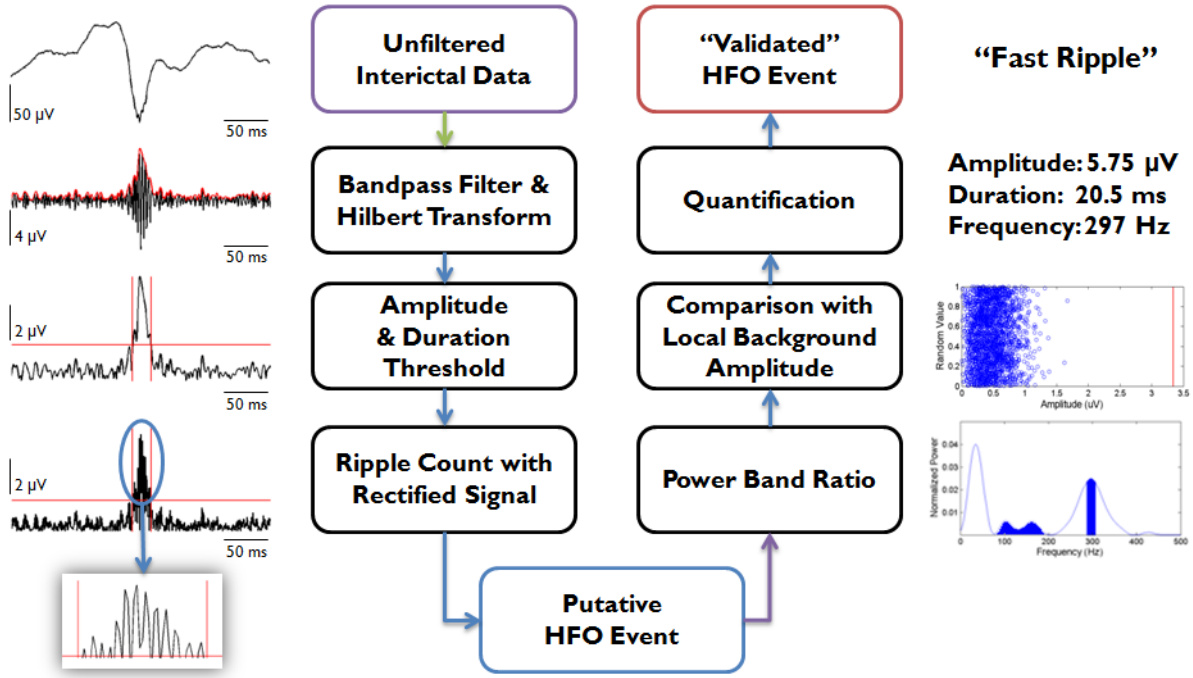
Zijlmans M, Jiruska P, Zelmann R, Leijten FS, Jefferys JG, Gotman J. High-frequency oscillations as a new biomarker in epilepsy. Ann Neurol 2012; 71: 169-178.

Supplementary Table

	first stage		second stage	
	ripples	fast ripples	ripples	fast ripples
accuracy	97.28%	93.46%	90.46%	84.20%
PPV	56.75%	32.23%	90.29%	81.32%
true positive	500	343	465	309
false positive	381	721	50	71
true negative	14	24		

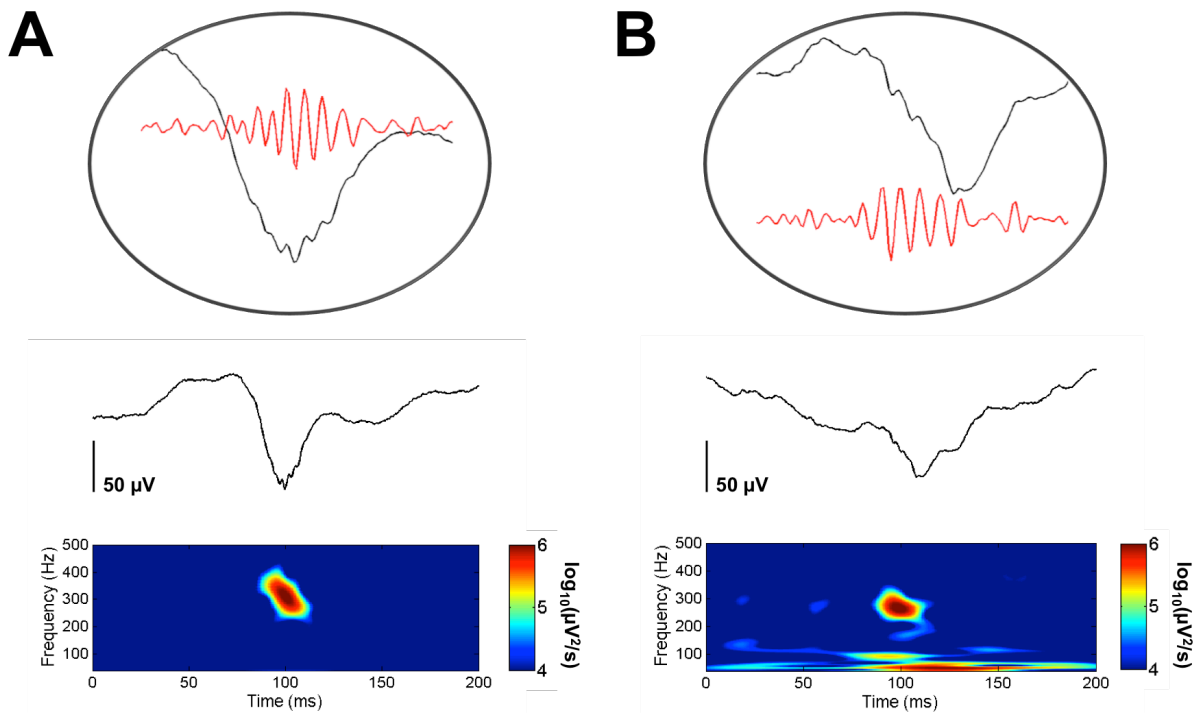
Supplementary Table 1: HFO detector performance.

Supplementary Figure 1



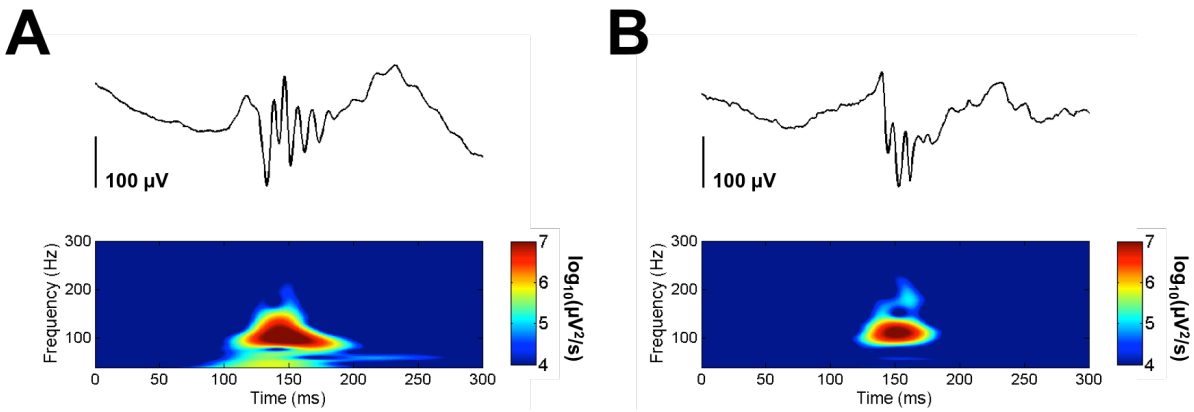
Supplementary Figure 1: Data processing schematics for automatically detecting HFO. Unfiltered interictal data were first band-pass filtered to either FR (200~500 Hz) or R (60~200 Hz) range and Hilbert-transformed. Instantaneous amplitude was obtained by taking the absolute value of Hilbert-transformed signal. Putative HFO events were selected if local maxima of the signal envelope exceed the predetermined amplitude and duration thresholds. They had to contain at least 6 peaks in the rectified signal to pass ripple count criterion. To reject false detections, the power band ratio was calculated and the mean amplitude of selected events was compared to local background amplitude (see materials and methods for details). Then, the amplitude, duration and peak frequency of appropriately detected events were quantified.

Supplementary Figure 2



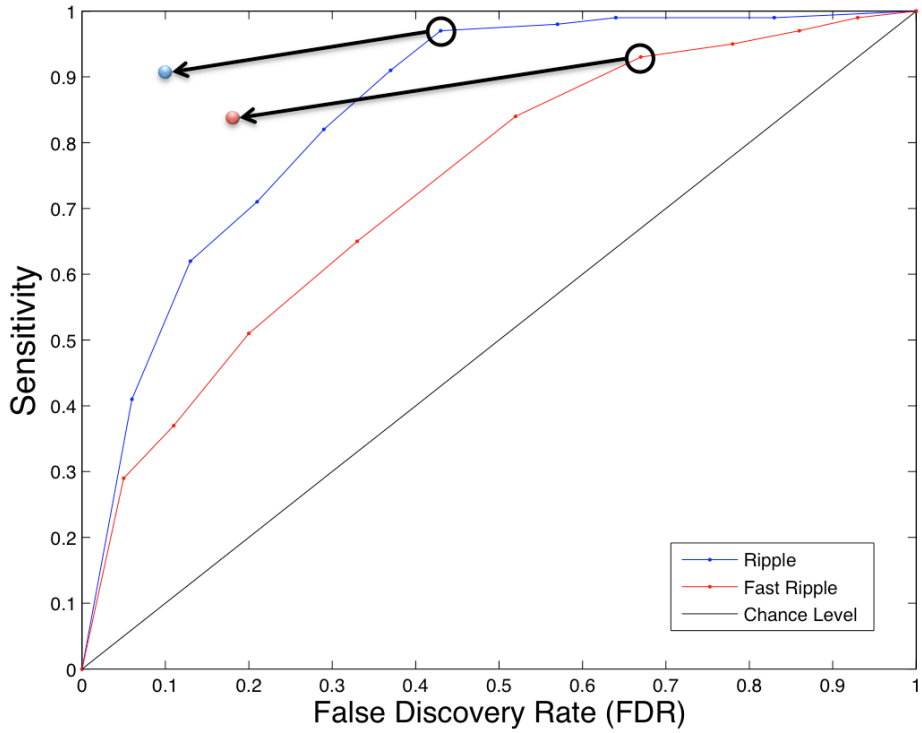
Supplementary Figure 2: Representative examples of detected FR events. Raw signal is shown in the middle. In the top panel within circles, magnified raw (black line) and filtered (red line, 10X, filtered to 200~500 Hz) signals are illustrated. In the bottom, Morlet wavelet time-frequency decomposition plots are shown for each event. (A) FR event in superior temporal gyrus of patient 10. (B) FR in fusiform gyrus of patient 6.

Supplementary Figure 3



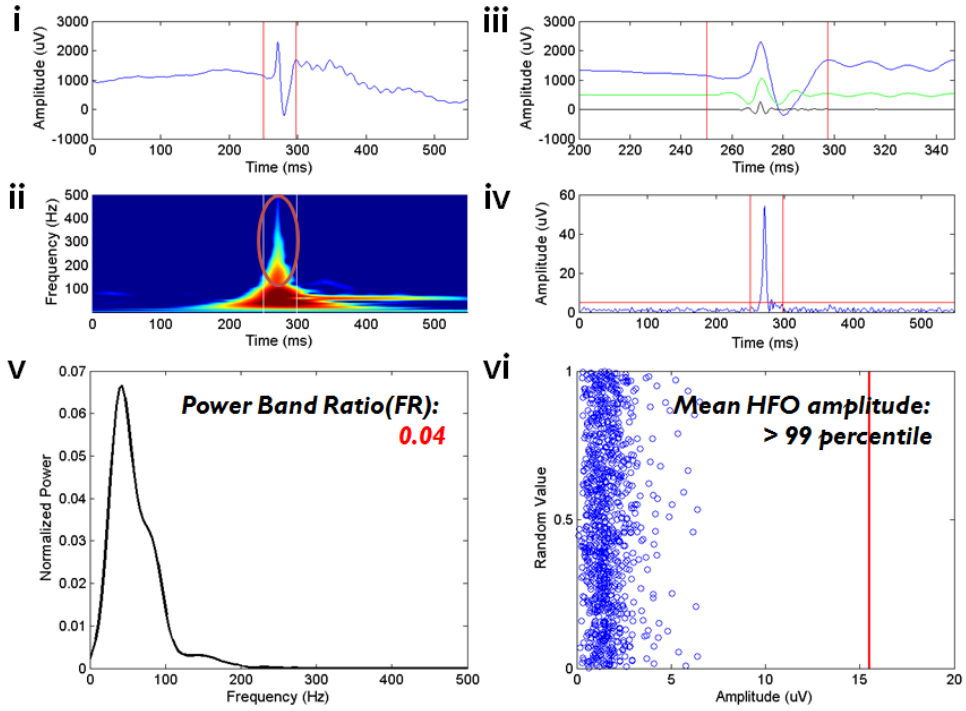
Supplementary Figure 3: Representative examples of detected R events. Raw signal is shown in the top panels. In the bottom, Morlet wavelet time-frequency decomposition plots are shown for each event. (A) R event in inferior temporal gyrus of patient 8. (B) R in inferior parietal lobule of patient 4.

Supplementary Figure 4



Supplementary Figure 4: ROC curve with different amplitude thresholds (from SD of 0.5 to 5.5 with step size of 0.5) across all tested data. Notice that this ROC curve is calculated based on sensitivity and false discovery rate, while traditional ROC curve is illustrated with sensitivity and specificity. Specificity measure requires the true negative events, which is problematic in rare event detection problems, such as in HFO detection. Blue curve is for R and Red for FR, while black straight line indicates the chance level detection. Black circles indicate the threshold of our choice for the first-stage detection, the SD of 3. Effects of second-stage post-processing steps are shown as blue and red filled circles. Notice the post-processing steps greatly reduce the false discovery rate at the cost of relatively small decrease in sensitivity.

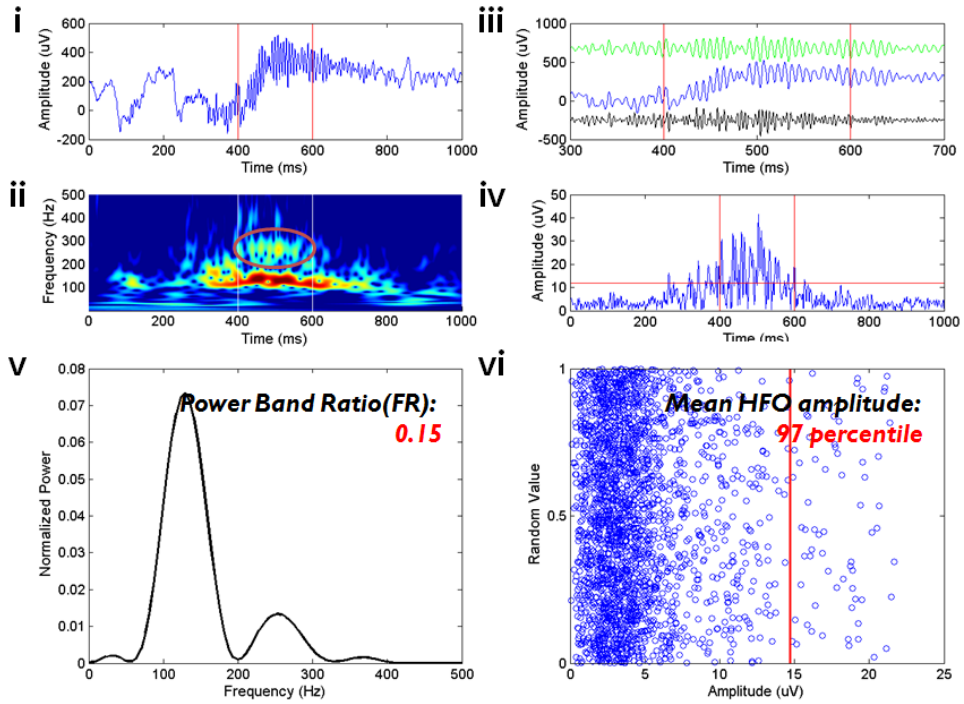
Supplementary Figure 5



Supplementary Figure 5: Type 1 false positive detections. (A) An example of high-frequency transients. Filtering of sharp and steep transients (including interictal epileptiform discharges) may generate "false" oscillations. This type of artifact causes broadband power increase. Waveform in this example was detected as candidate FR event. (i) Unfiltered data. (ii) Time-frequency decomposition plot. Sharp transient causes “candle-like” broadband power increase, although dominated in lower-frequency. (iii) Unfiltered signal (blue), R range filtered signal (green) and FR band filtered signal (red) at higher temporal resolution. (iv) FR-filtered amplitude. During event, the amplitude clearly exceeds amplitude and duration thresholds. (v) Calculation of power band ratio. The value was 0.04, which did not pass the criterion. (vi) Comparison with local background. The mean amplitude during detected events was larger than 99 percentile of local background amplitude.

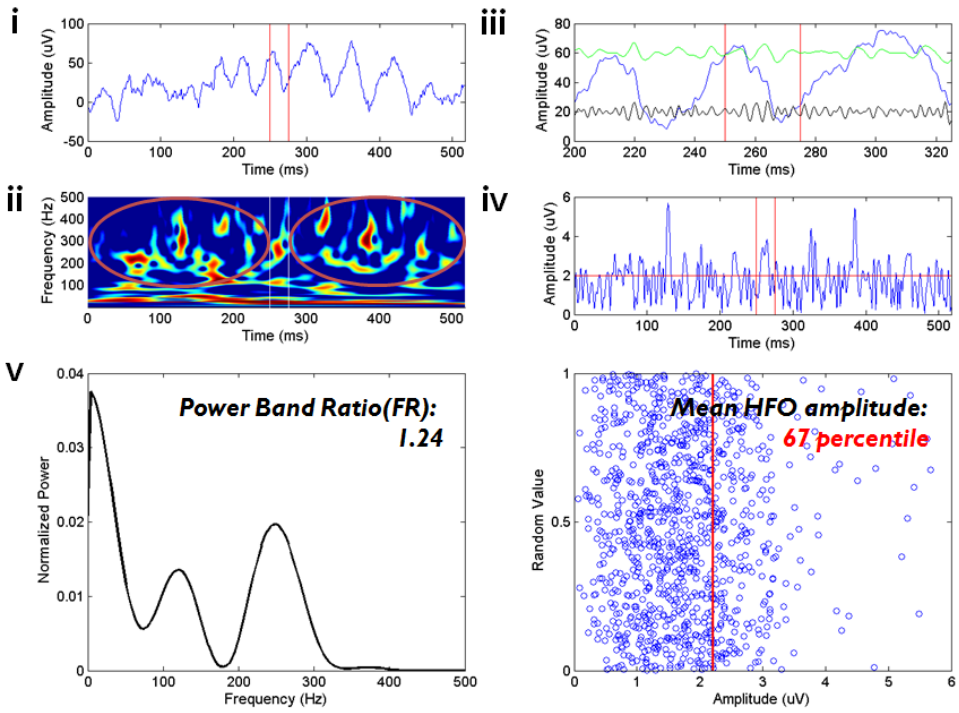
.

Supplementary Figure 6



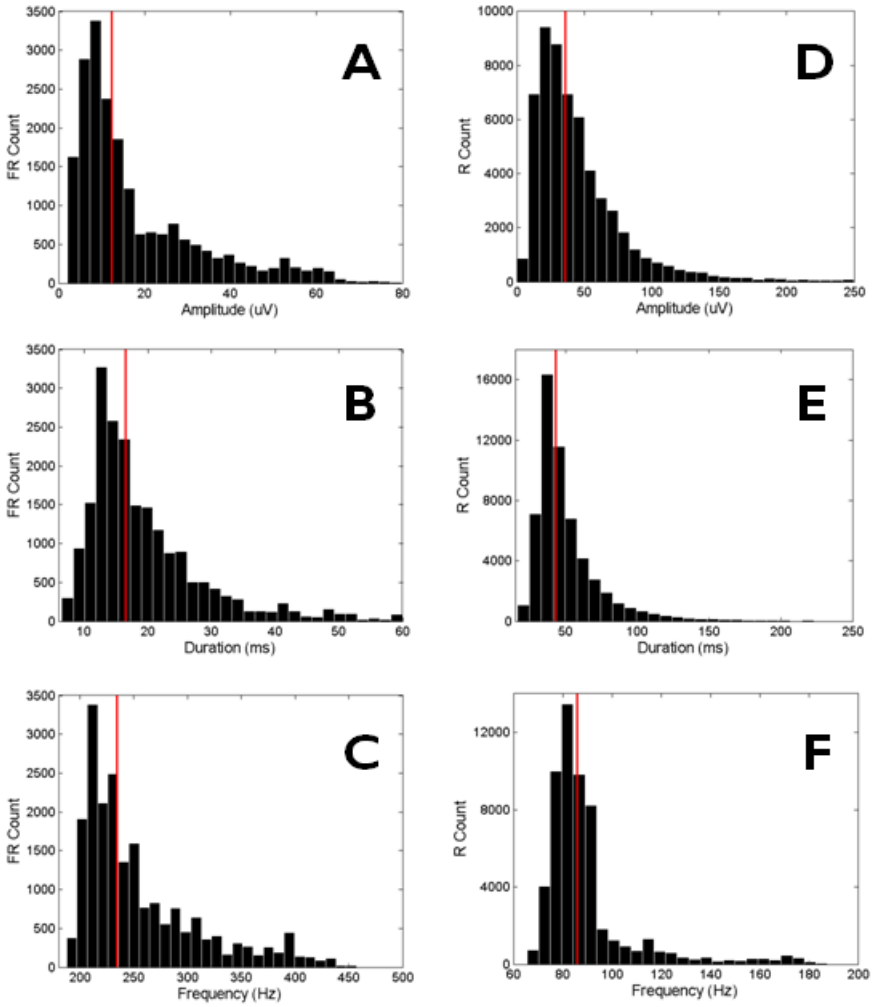
Supplementary Figure 6: Type 2 false positive detections. An example of lower-frequency harmonic event. This is a valid R event, but also detected as a candidate FR event due to harmonic component in the FR range. This event failed to pass both the power band ratio and local background amplitude comparison criteria. Panel organization is analogous to Supplementary figure 2.

Supplementary Figure 7



Supplementary Figure 7: Type 3 false positive detections. An example of transient high-frequency events with amplitudes larger than the global threshold but not discriminative from the local background. Although the power band ratio value exceeds 1, the mean amplitude of detected event corresponds only to 67th percentile of the local background amplitude. Panel organization is analogous to Supplementary figure 2. Notice comparable high-frequency blobs preceding and following the event in the time-frequency map, indicating noisy background activities.

Supplementary Figure 8



Supplementary Figure 8: Quantification results of all detected HFO in histograms. Red vertical lines indicate median values. (A) Amplitude of FR. (B) Duration of FR. (C) Peak frequency of FR. (D) Amplitude of R. (E) Duration of R. (F) Peak frequency of R.